AD_____

GRANT NUMBER DAMD17-96-1-6226

TITLE: Computer-Aided Diagnosis of Breast Cancer: A Multi-Center Demonstrator

PRINCIPAL INVESTIGATOR: Carey E. Floyd, Jr., Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: October 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1997	Annual (1 Oct 96 - 30 Sep 97)

4. TITLE AND SUBTITLE Computer-Aided Diagnosis of Breast Cancer: A Multi-Center Demonstrator

5. FUNDING NUMBERS
DAMD17-96-1-6226

6. AUTHOR(S)
Carey E. Floyd, Jr., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University Medical Center
Durham, North Carolina 27710

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a computer aided diagnosis (CAD) system for early prediction of breast cancer using the BI-RADS™ findings reporting criteria provided by mammographers distributed over a wide geographical area.

In the first year of this project, we have hired a Data Technician to set up and manage the mammographic findings database. So far, 700 hundred cases from Duke have been entered, as well as 1000 from the University of Pennsylvania. A further 500 cases from Sloan-Kettering Cancer Center are being processed and entered, as well as an additional 100 from Duke. The data collection process has been delayed by the decreased budget, which in turn delayed CAD system testing. An artificial neural network (ANN) to predict biopsy outcome has been developed. A genetic algorithm has been developed for selecting subsets from the dataset in order to decrease cross-validation variance and increase the network's performance in the ROC area.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	13
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Carey E. Floyd 8 April 98

PI - Signature Date

Table of Contents

Progress report for DAMD17-96-1-6226, Computer Aided Diagnosis of Breast Cancer: A Multi-Center Demonstration.

PI: Carey E. Floyd Jr.

Abstract

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a computer aided diagnosis (CAD) system for early prediction of breast cancer using the BI-RADS™ findings reporting criteria provided by mammographers distributed over a wide geographical area.

In the first year of this project, we have hired a Data Technician to set up and manage the mammographic findings database. So far, 700 hundred cases from Duke have been entered, as well as 1000 from the University of Pennsylvania. A further 500 cases from Sloan-Kettering Cancer Center are being processed and entered, as well as an additional 100 from Duke. The data collection process has been delayed by the decreased budget, which in turn delayed CAD system testing. An artificial neural network (ANN) to predict biopsy outcome has been developed. A genetic algorithm has been developed for selecting subsets from the dataset in order to decrease cross-validation variance and increase the network's performance in the ROC area.

Introduction

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from patients' mammographic findings and medical history. Specifically, this system will predict the malignancy of non-palpable lesions that are examined with diagnostic mammography and are considered for biopsy.

The lifetime risk of developing breast cancer has increased steadily from 1940, when the first statistics were collected, to the present risk of one woman in eight (Garfinkel, Boring et al. 1994). Several large studies have demonstrated that screening mammography results in an approximately 30% decrease in mortality due to breast cancer (Verbeek, Hendriks et al. 1984; Shapiro 1994). Unfortunately, evaluating mammograms is a complicated task. To determine whether a lesion is benign or whether further action such as close follow-up or biopsy for histologic diagnosis is warranted, multiple radiographic features of each mammographic abnormality must be considered in combination with the patient's age, history and physical exam.

Although mammography is a sensitive tool for detecting breast cancer, the positive predictive value (PPV) has historically been low (Ciatto, Cataliotti et al. 1987; Adler and Helvie 1992; Kopans 1992). Due to several factors, including overlap of the radiographic appearance of benign and malignant breast lesions (Ciatto, Cataliotti et al. 1987) as well as an overall conservative approach of physicians (Hall 1986), only 10-34% of women undergoing biopsy for mammographically suspicious nonpalpable lesions have a malignancy by histologic diagnosis (Kopans 1992). This relatively low PPV of mammography-induced biopsy raises several problems. If the mammography screening recommendations of the American College of Radiology (ACR) and the American Cancer Society (ACS) are fully implemented, nearly all women over the age of 50 would undergo a yearly mammogram. Continuing today's biopsy rate in the range of 0.5 - 2.0% of mammographic exams could result in over one million biopsies performed each year (Hall,

Storella et al. 1988). Clearly, due to the present low PPV of mammography, hundreds of thousands of women undergoing biopsy for a benign finding would be unnecessarily subjected to the discomfort, expense, potential complications, change in cosmetic appearance, and anxiety that can accompany breast biopsy (Helvie, Ikeda et al. 1991; Dixon and John 1992; Kopans 1992; Schwartz, Carter et al. 1994). Moreover, the financial burden of these procedures could well be unacceptable in the present political and economic climate to reduce expenditures (Hall, Storella et al. 1988; Kopans 1992; Schwartz, Carter et al. 1994).

In order to improve the PPV and specificity of film-screen mammography, an artificial neural network (ANN) has been constructed to assist radiologists in the differentiation of benign from malignant lesions. Inputs to the ANN are derived from the patient's history and the radiologist's description of lesion morphology following the ACR Breast Imaging Reporting and Data System (BI-RADS^{TM}). The output of the neural network is the likelihood of malignancy. This ANN will provide an accurate prediction of malignancy for the physician to consider when contemplating the decision to biopsy.

The development of this system will provide three significant improvements for early breast cancer detection: 1)increase the diagnostic accuracy of mammography for predicting malignancy of breast lesions; 2)decrease the number of patients sent to biopsy with benign lesions (and thus provide a significant savings of healthcare costs); and 3)decrease the variability of diagnosis for mammography. This last will be a result of the development of a computer algorithm since it has no intra-observer variability.

Toward this goal, we have developed an artificial neural network (ANN) to predict biopsy outcome from mammographic and history findings. In the first year of the grant we have 1) developed a database for indexing and manipulating mammographic findings, 2) acquired 200 new cases from Duke using the standardized BI-RADS^{TM} reporting system, 3) acquired 1000 cases from the University of Pennsylvania, 4)formed agreements with two other hospitals to acquire data.

The goal of this work is to improve the specificity of diagnosis with little loss of sensitivity thus significantly improving the positive predictive value of breast biopsy. In this demonstration project, we proposed to acquire cases from other institutions to evaluate how well the model can translate to other patient populations and other radiologists readings.

What follows is a point by point assessment of the progress for each task in the statement of work:

Statement of Work

(months 1-36)

1) Acquire diagnostic mammography cases from mammography providers distributed over a wide geographical area using the BI-RADS™ findings reporting criteria.

(months 1-6) Develop tools for managing the database and generating reports)

Cases will be acquired from each site and entered into the database as a continual effort.

(months 1-36)

2) Test the existing CAD system on biopsy cases from other mammographic facilities (external to Duke). This testing will be performed on a monthly schedule. The results will be summarized at the end of the first six months and periodically through the project.

3) Develop an ANN to predict biopsy outcome from BI-RADS™ mammographic and history findings for the individual and combined datasets from other mammographic facilities.

(months 1-6) Develop tools for importing cases from the database into the artificial neural network systems.

(months 6-12) Refine the coding of the ANNs to facilitate use with large datasets.

(months 6 -36) Examine the behavior of the different training techniques: cross-validation, bootstrap, and round robin as the datasets grow in size.

4) Evaluate the difference between the individual and combined networks.

(months 6-36) This work will begin in the first year as the data and tools become available. It will continue throughout the project.

****

Progress in the first period (months 1 - 12)

(months 1-36)

1) Acquire diagnostic mammography cases from mammography providers distributed over a wide geographical area using the BI-RADS™ findings reporting criteria.

(months 1-6) Develop tools for managing the database and generating reports)

In the first year of this project we have:

1 Hired a Data Technician to set up and manage the database.

2 Implemented a database in the FOXPRO database language.

3 Entered the existing 700 cases from Duke into the database.

Cases will be acquired from each site and entered into the database as a continual effort.

4 Acquired 1000 cases from the University of Pennsylvania. These cases were converted to our feature scoring scheme and then entered into the database. Tools were developed to compare the distributions of findings for the U Penn cases to those from Duke.

5 Negotiated with the University of Virginia and Sloan-Kettering Cancer Center to obtain 1000 cases from each in the next year of the grant. Agreement is almost complete.

6 Discussed data acquisition with University of San Francisco and regretfully was told that they would not be participating due to the decrease in our payment scale per case (necessitated by the negotiated budget reduction of a factor of two).

7 After discussion with our potential collaborators, we realized that prospective acquisition of cases would not be practical given the budget cuts. Therefore, we agreed to accept retrospective data as long as it was acquired using the BI-RADS™ findings reporting criteria. This has no effect on the scientific aims of the study. It does however, simplify the data acquisition for both our collaborators and ourselves.

(months 1-36)

2) Test the existing CAD system on biopsy cases from other mammographic facilities (external to Duke). This testing will be performed on a monthly schedule. The results will be summarized at the end of the first six months and periodically through the project.

The Penn data did not arrive until month 10. The monthly acquisition schedule has been revised as described in [7] above. We have begun the comparison and have discovered an important sampling issue:

8 Begun to examine the effects of sampling strategies on networks that include both the Duke and Penn data.

9 Explored the use of a genetic algorithm to optimize preprocessing of the findings (from each of the two sets of data) for the network models.

10 Developed a genetic algorithm for selecting subsamples from the data set such that the distribution of inputs and outcomes will be similar between the sets. This was motivated by the observation that the performance of a network will be significantly affected by the ratio of cases with masses to cases with calcifications. A network will

perform with ROC are in the 90's for a dataset with only masses while it will only perform in the 70's for a set with just calcifications. In a cross validation technique, the dataset is divided into several samples of equal size and the training and testing sets are formed from this partitioning. If each partition has an imbalance of mass and calcification cases, the performance will vary considerably over the different training/testing sets. To improve the consistency of the cross-validation technique, we propose to use this genetic algorithm to construct cross-validation partitions that have uniform distributions of the findings. The effect of this technique on the cross-validation variance will be investigated in the second year.

3) Develop an ANN to predict biopsy outcome from BI-RADS™ mammographic and history findings for the individual and combined datasets from other mammographic facilities.

(months 1-6) Develop tools for importing cases from the database into the artificial neural network systems.

Done

(months 6-12) Refine the coding of the ANNs to facilitate use with large datasets.

(months 6 -36) Examine the behavior of the different training techniques: cross-validation, bootstrap, and round robin as the datasets grow in size.

4) Evaluate the difference between the individual and combined networks.

(months 6-36) This work will begin in the first year as the data and tools become available. It will continue throughout the project.

Underway, see [8,9,10] above.

The original goal was to obtain about 1500 cases per year (external to Duke). Since the budget was cut by a factor of two, this goal will be a challenge to meet. With the agreements we hope to finalize soon, we will have 3000 cases by the end of year two which will put us on target.

Publication

In the current period, we have not published any manuscripts describing work funded in whole or in part by this grant. We have begun writing several manuscripts that will be submitted in the next year.

Conclusion

In conclusion, the data acquisition efforts are underway with significant progress already made. The analysis of the performance of the model on the new data will be completed in the second year. The project is on schedule, however there are no scientific results to report at this time. This status is in keeping with the statement of work.

References

Adler, D. D. and M. A. Helvie (1992). "Mammographic biopsy recommendations." Current Opinion in Radiology **4**: 123-129.

Ciatto, S., L. Cataliotti, et al. (1987). "Nonpalpable lesions detected with mammography: review of 512 consecutive cases." Radiology **165**: 99-102.

Dixon, J. M. and T. G. John (1992). "Morbidity after breast biopsy for benign disease in a screened population." Lancet **1**: 128.

Garfinkel, L., C. C. Boring, et al. (1994). "Changing trends: an overview of breast cancer incidence and mortality." Cancer **74**(1): 222-227.

Hall, F. M. (1986). "Screening mammography - potential problems on the horizon." New England Journal of Medicine **314**: 53-55.

Hall, F. M., J. M. Storella, et al. (1988). "Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography." Radiology **167**: 353-358.

Helvie, M. A., D. M. Ikeda, et al. (1991). "Localization and needle aspiration of breast lesions: complications in 370 cases." AJR. American Journal of Roentgenology **157**: 711-714.

Kopans, D. B. (1992). "The positive predictive value of mammography." AJR. American Journal of Roentgenology **158**: 521-526.

Schwartz, G. F., D. L. Carter, et al. (1994). "Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential." Cancer **73**: 1660-1665.

Shapiro, S. (1994). "Screening: assessment of current studies." Cancer **74**: 231-238.

Verbeek, A. L. M., J. H. C. L. Hendriks, et al. (1984). "Reduction of breast cancer mortality through mass screening with modern mammography." Lancet **1**: 1222-1224.